ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Review of Machine Learning based Sentiment Analysis on Social Web Data

Monelli Ayyavaraiah

Assistant Professor, Dept. of IT, Mahatma Gandhi Institute of Technology, Hyderabad-500075, TS, India

ABSTRACT: Sentiment analysis or opinion mining is the information retrieval strategy that delivers the vision of the relevant users such as customers about entities such as services or products, individuals such as service providers or sellers, functional issues involved, events and their attributes. As an example service provision or product selling with monitorial or other interests, intended to morph according to target users vision. In other dimension intended users seeks the opinion of the existing users, which is in order to select productive service provider among the multiple providers available. Hence the opinion mining or sentimental analysis is critical factor and challenging. In the era of social web that includes social networks, forums and blogs, the sentiment analysis in order to notice the public opinion is critical in decision making activities by an individual or an organization. The phenomenal growth in the data quantity of social web, the manually analyzing the opinion is almost impractical. Hence the machine based sentiment analysis or opinion mining is desired. In this context the automated sentiment analysis on social web become critical research objective that grabbed researcher's attention over a decade. In this article we carried out the review of the automated sentiment analysis that addressed contemporary affirmation of the existing literature and the future scope of the research related to this objective.

KEYWORDS: Opinion Mining, Sentiment Analysis, Social Web Data, Machine Learning, Social Media

I . INTRODUCTION

Sentiment analysis or opinion mining is machines analyzing human expressions of sentiment. Human according to various thoughts, actions, or reactions generate feelings of subjective nature such as emotion, mood, combined with visible facial expressions or postures, and communicate using language either in the spoken or written form. Opinions expressed by others are a matter of interest for everyone be it individuals or companies. Individuals through reviews, blogs, and opinions expressed on social media by other people, buy a product, or follow the popularity of various political parties to cast their vote. This plethora of information comprising of peoples thoughts, likes, dislikes shared among different related and unrelated people determines to a large extent other individuals
choices and preferences in liking or buying a product or in supporting representatives of political parties. Companies deeply mine consumer reviews for brand management and for promoting their products [1]. In economics and finance to understand beyond fundamental and technical knowledge analysis, sentiment analysis supporters suggest additionally it is essential to use information as diverse as, impending announcements, sudden surge in commodity prices, rumors and reports of a market collapse or break through, increase in the interest rates by central banks, fluctuations in dollar prices, etc. as these factors help in better estimating and forecasting situations of changes in market. In Fig 1 the process flow of opinion mining and sentiment analysis is shown.

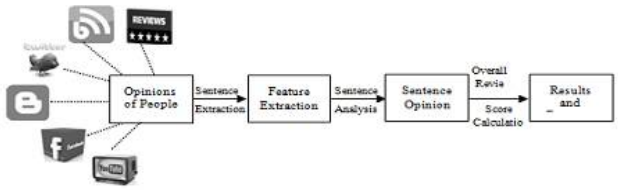

Figure 1: The architecture of Sentiment Analysis or Opinion Mining

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

A key part of the communication process is expressing or signifying the event with descriptions of emotion, mood, and sentiment. An affect is expressed with text and speech combined with fitting descriptions and this language specific knowledge is crucial for sentiment analysis. In the book "Understanding figurative language" by Sam Glucksberg [2] the metaphors "my spouse's lawyer is a shark and my job is my jail" characteristically expresses an effect that is suggestively different besides the outward. The descriptions [3] for expressing an effect, are generally used in everyday language, are popular in the field of fiction content, extend generally to scientific languages [4] are more predominant in natural [5] and biological sciences [6] and to a certain extent used also in the domains of finance and economics. A sentiment analysis systems objective or purpose is to detect and tabulate an author/speaker opinion, or find how discerning the author/speaker is in giving value how much or how little to an object, event, state, or abstract idea. To handle the notions of quantitative semantics spatial relationships are applied using cognitive capabilities of the computer emulating the human mind [7]. The objective of performing sentiment analysis finally is detecting if any reader/listener might have an instinctive emotional response for a speaker/author and to tabulate the responses for further analysis.

II. CONTEMPORARY LITERATURE OF MACHINE LEARNING BASED SENTIMENT ANALYSIS

The techniques of machine learning have found wide applications in sentiment analysis [8][9][10][11][12][13]. The machine learning model is applied with the reduced and optimal feature vector obtained after feature selection. The most popular methods that are widely used are, naive Bayes (NB), support vector machine (SVM), artificial neural networks (ANN), and maximum entropy (MaxEnt) [14], for almost all the sentiment analysis issues [15][12][16][17][18]. Pang et al. [12] experimented sentiment analysis with a movie review dataset using different

ML algorithms like SVM, NB, and MaxEnt. In the sentiment analysis tests SVM outperformed the remaining ML methods. A similar investigation by Tan and Zhang [17] is implemented with SVM and other classifiers. In this work the authors state the performance of SVM algorithm for sentiment analysis is better compared to others. O'Keefe and Koprinska [19] explored sentiment analysis by applying NB and SVM classifiers with different feature weighting and feature selection methods. In the tests conducted SVM classifier performance is shown to be better compared to the NB classifier. Ye et al. [20] explored 3 supervised learning algorithms with travel destination domain reviews dataset. The algorithms SVM, NB, and character-based N-gram model are used, and frequency of words is applied for assigning feature weights. In the sentiment analysis experiments the results demonstrate SVM outperforming the remaining classifiers. Cui et al. [21] experimented with sentiment analysis using discriminative classifiers like, SVM, winnow classifier, and other generative models, and state SVM is more suitable compared to the other methods. Moraes et al. [15] empirical study of document-level sentiment analysis with the classifiers SVM and ANN compares them with their limitations. Saleh et al. [22] tested various domains of datasets for sentiment analysis with SVM classifier and different weighting methods. Li et al. [23] experimented by constructing several classifiers using dissimilar feature sets like unigrams and the part of speech based features. Next they combined these classifiers based on different combination rules and tested for their performance.

The test results show the combined classifier outperformed the individual classifiers performance. Tsutsumi et al. [24] devised and explored an integrated classifier that combines maximum entropy, SVM, and classifiers based on score calculation with a movie review dataset and the results show better classification performance. Osajima et al. [25] devised approach for a dataset of review documents has the sentences classified based on their polarity by using word polarity aggregation. In [26] a three-phase framework is proposed by the authors to select optimal combination of the classifiers for the assembly of multiple classifiers for sentiment analysis. Xia et al. [27] devised several forms of ensemble techniques for different features categories such as, based on Part of Speech, word relation based, etc. and with different classifiers like NB, SVM, MaxEnt for investigating the sentiment analysis performances. Prabowo and Thelwall [28] combined three methods, machine learning, rule based classification, and supervised learning to devise a hybrid classifier for enhancing the effectiveness of the classification. Dinu and Iuga [29] performed experiments with applications of opinion mining using the naive Bayes classifier. Kang et al. [30] devised a scheme for improving the classifiers NB and SVM.

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

The proposed approach is tested with a restaurant reviews dataset.In machine learning numerous models have been devised that find solve the problems of sentiment analysis, and the most widely used methods are, SVM, Naive Bayes, maximum entropy, hidden Markov models and decision trees. Among these methods SVM has been regarded for sentiment analysis as the best algorithm of machine learning. However recently Agarwal and Mittal [31] demonstrated, the BMNB classifier using mRMR feature selection scheme for sentiment analysis could outperform SVM classifier. They stated that by the inclusion of features that are more informative and less redundant, improvement in the machine learning models performance could be achieved.

III. IMPACT OF MACHINE LEARNING IMPACT ON SENTIMENT ANALYSIS

Most of the contemporary models are based on machine learning models such as

1) Naive Bays Classifier.
2) Support Vector Machine (SVM).
3) Multilayer Perceptron.
4) Clustering.

The significance and limits of these machine learning techniques learned are:

- There are a number of benefits with Naïve Bays Classifier. Two major benefits are, it is simple to implement, and is effective in computation. However the major drawback with the approach is that it assumes attributes based on probability and as a result useless attributes are generated more.
- The benefits of the models built using SVM classifier have been determined to be highly accurate in prediction, and effective in solving the problems of dimensionality. However it is a highly complex model to implement in case missing values are present in a dataset.
- The benefits of Multilayer Perception are, it performs as a universal function approximation, and is capable of creating strong relations among the variables of input and output based on the strategy of learning and building. However the method incurs high overhead and also in the implementation a dense training set is needed.
- The benefit of clustering approach is that it assures optimally high decision making capabilities by generating multiple classes. The disadvantages are that as there is no training included in the implementation the classes cannot be known beforehand, and a high number of classes are generated leading to many complications. Also the model used for measuring features and distance mostly determines its applications.

IV. FUTURE RESEARCH OBJECTIVES

In the process of sentiment analysis several topics of open research exist. A few of them are as follows:

- The modeling of the compositional sentiment has to be further improved. It implies better computation efficiency of the sentence sentiment related to sentiment shifters, sentence structure, and sentiment-bearing words.
- The question of automatic entity resolution that is denoting numerous names to the same product inside and across documents has to be addressed. The application of anaphora resolution with efficiency is also a most important issue that has to be solved. The issue of aspect extraction technique for grouping aspects another difficult that is. E.g. To talk about a phone, terms such as “battery life” or “power usage” that denote one aspect create number of difficulties that has to be answered.
- In detecting for every entity relevant text, where many entities may be discoursed in a document, the existing techniques accuracy is of insufficient levels that needs improvement.
- In the detection of sarcasm there are a few methods of classification used however these techniques have to be built into systems of autonomous sentiment analysis.
- A major problem of many systems of sentiment analysis is handling noisy texts involving mistakes of spelling/ grammar, punctuation are missing/unpredictable and use of slang.

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

- The existing methods of sentiment analysis are designed to find subjective statements sentiment and not that of objective statements that regularly show up in news articles and though of factual type they however also hold sentiment. The objective statements have to be associated with sentiment scores by the context based algorithms.
- The integration of sentiment analysis with the latest methods of soft computing and machine learning is required and has to be an important part of future research studies. These methods and their strategies have majorly risen in popularity due to their contributions in recent times and need to be further researched for enhancing sentiment analysis systems.
- The depiction of the content in the form of metaphors is lexically highly challenging in sentiment analysis. This necessitates a great deal of research in the area of feature extraction and optimality detection.

V. CONCLUSION

This manuscript contributed a systematic review of sentiment analysis and opinion mining. The complexity of data presentation and dimensionality, diversified usage requirements, the sentiment analysis or opinion mining emerged as critical research objective since a decade. This review explored the sentiment analysis process, contemporary review of the machine learning based sentiment analysis models found in recent literature, impact of machine learning on sentiment analysis and possible and potential research objectives for future research. Finally, we conclude the manuscript by saying that all the sentiment analysis tasks are very challenging, since understanding and knowledge of the problem and its solutions are still limited. The main reason is that it is a natural language processing task, and natural language processing has no easy problems. However, many significant progresses have been made. Finally it is obvious to conclude that the sentiment analysis is having potential scope for future research and one of that is exposing the scope of evolutionary computational or soft computing techniques and the hybridizing these techniques towards feature extraction, selection to classify the sentiment.

REFERENCES

[1] Candillier, Laurent, Frank Meyer, and M. Boullé., " Comparing state-of-the-art collaborative filtering systems. International conference on Machine Learning and Data Mining MLDM ", Leipzig/Germany. Lecture Notes in Computer Science, Volume 4571:548– 562. Berlin: Springer,2007.
[2] Glucksberg, Sam., "Understanding figurative language: From descriptions to idioms" ,Oxford: Oxford University Press,2001.
[3] Goatly, Andrew., "The language of descriptions", London: Routledge,1997.
[4] Miller, A.L., "Imagery in scientific thought. Boston, Basel ", Stuttgart: Birkhaeuser,1984.
[5] Pullman, B., " The atom in the history of human thought", Oxford: Oxford University Press,1988.
[6] Verschuuren, G.M., " Investigating the life sciences: An introduction to the philosophy of science", Oxford: Pergamon Press,1986.
[7] Hobbs, Jerry R Literature and cognition," Lecture notes, number 21, Center for the Study of Language and Information ", Stanford, California,1990.
[8] Agarwal B, Mittal N.," Optimal feature selection for sentiment analysis", In: Proceedings of the 14th international conference on intelligent text processing and computational linguistics (CICLing 2013), vol 7817, no 1, Samos, pp 13–24 ,2013.
[9] Agarwal B, Mittal N, Cambria E., " Enhancing sentiment classification performance using bi-tagged phrases", In: Proceedings of the 13th IEEE international conference on data mining workshops, Dallas, pp 892–895 ,2013.
[10] Chikersal P, Poria S, Cambria E.," SeNTU: sentiment analysis of tweets by combining a rule-based classifier with supervised learning", In: Proceedings of the international workshop on semantic evaluation, (SemEval 2015), Denver ,2015.
[11] Prerna C, Poria S, Cambria E, Gelbukh A, Siong CE.," Modelling public sentiment in Twitter: using linguistic patterns to enhance supervised learning", In: Computational linguistics and intelligent text processing. Springer International Publishing, Cham, pp 49– 65,2015.
[12] Pang B, Lee L, Vaithyanathan S0., "Thumbs up? Sentiment classification using machine learning techniques", In: Proceedings of the conference on empirical methods in natural language processing (EMNLP), Prague, pp 79–86 ,2002.
[13] Poria S, Gelbukh A, Cambria E, Das D, Bandyopadhyay S " Enriching SenticNet polarity scores through semi-supervised fuzzy clustering", In: 2012 IEEE 12th international conference on data mining workshops (ICDMW), Brussels, pp 709–716 ,2012.
[14] Witten IH, Frank E, Hall MA., " Data mining: practical machine learning tools and techniques", 3rd edn. Morgan Kaufmann, Burlington,2011.
[15] Moraes R, Valiati JF, Neto WPG., "Document-level sentiment classification: an empirical comparison between SVM and ANN", Expert Syst Appl 40(2):621–633,2013.
[16] Pang B, Lee L., " Opinion mining and sentiment analysis. Foundations and trends in information retrieval", vol 2, no 1–2. Now Publishers, Hanover, pp 1–135,2008.
[17] Tan S, Zhang J.," An empirical study of sentiment analysis for chinese documents", Expert Syst Appl 34(4):2622–2629 ,2008.
[18] Zhu J, Xu C, Wang HS.," Sentiment classification using the theory of ANNs" J China Univ Posts Telecommun 37(1):58–62,2010.

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

[19] O'keefe T, Koprinska I.," Feature selection and weighting methods in sentiment analysis", In: Proceedings of the 14th Australasian document computing symposium, Sydney, pp 67–74,2009.
[20] Ye Q, Zhang Z, Law R., " Sentiment classification of online reviews to travel destinations by supervised machine learning approaches",. Expert Syst Appl 36(3):6527–6535,2009.
[21] Cui H, Mittal V, Datar M .," Comparative experiments on sentiment classification for online product reviews", In: Proceedings of the 21st national conference on artificial intelligence, Boston, pp 1265–1270,2006.
[22] Saleh MR, Martin-Valdivia MT, Montejo-Raez A, Urena-Lopez LA., " Experiments with SVM to classify opinions in different domains", Expert Syst Appl 38(12):14799–14804,2011.
[23] Li S, Zong C, Wang X Sentiment classification through combining classifiers with multiple feature sets. In: Proceedings of the international conference on natural language processing and knowledge engineering (NLP-KE), Beijing, pp 135–140 ,2007.
[24] Tsutsumi K, Shimada K, Endo T.," Movie review classification based on a multiple classifier", In: Proceedings of the annual meetings of the Pacific Asia conference on language, information and computation (PACLIC), pp 481–488,2007.
[25] Osajima I, Shimada K, Endo T.," Classification of evaluative sentences using sequential patterns", In: Proceedings of the 11th annual meeting of the association for natural language processing, Takamatsu, pp 1–8 ,2005.
[26] Lin Y, Wang X, Zhang J, Zhou A., "Assembling the optimal sentiment classifiers", In: Proceedings of the 13th international conference on web information systems engineering, vol 7651, no 1, Paphos, pp 271–283,2012.
[27] Xia R, Zong C, Li S., " Ensemble of feature sets and classification algorithms for sentiment classification", J Inf Sci 181(6):1138–1152, 2011.
[28] Prabowo R, Thelwall M., " Sentiment analysis: a combined approach", J Informetr 3(2):143–157,2009.
[29] Dinu LP, Iuga I., "The Naive Bayes classifier in opinion mining: in search of the best feature set", In: Proceedings of the 13th international conference on intelligent text processing and computational linguistics, CICLing, vol 7181, no 1, New Delhi, pp 556–567,2012.
[30] Kang H, Yoo SJ, Han D., "Senti-lexicon and improved Naive Bayes algorithms for sentiment analysis of restaurant reviews", Expert Syst Appl 39(5):6000–6010 ,2012.
[31] Agarwal B, Mittal N., "Prominent feature extraction for review analysis: an empirical study", J Exp Theor Artif Intell. Taylor Francis. doi:10.1080/0952813X.2014.977830, 2014.

BIOGRAPHY

Monelli Ayyavaraiah received B.Tech degree in Information Technology from SV University in 2011. He received M.Tech degree in Computer Science and Engineering, from JNTUA in 2013. He is currently working as Asst.Professor at MGIT, Hyderabad. His current research interests are Web Mining, Social web and sentiment analysis and machine learning.